

PRESENTACIÓN DE PROPUESTA DEL PROYECTO

An abstract graphic element consisting of numerous thin, curved purple lines that curve and overlap to create a sense of depth and motion, resembling waves or a grid.

Jerónimo Vargas

Carlos Gómez

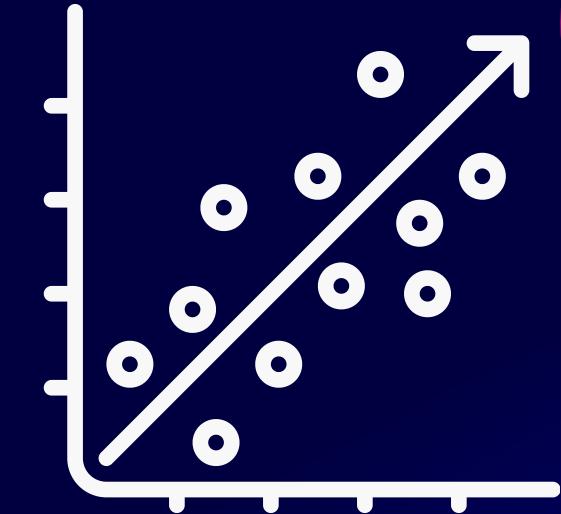
Juan Esteban López

Agenda

- 3 Elección del cliente y análisis del negocio
- 4 Supuestos del proyecto
- 5 Limpieza y alistamiento de los datos
- 8 Exploración de datos
- 14 Modelamiento
- 18 Tablero interactivo
- 22 Despliegue del tablero

CLIENTE

Área financiera del sistema:
Viabilidad financiera del
proyecto



COMO AYUDAR AL CLIENTE:

1. ¿Cuál es el costo que debe tener una bicicleta por hora para que el proyecto sea viable teniendo en cuenta los costos y la demanda?
2. ¿Cuál es el costo del sistema y como se distribuyen en costos de operación y costos de mantenimiento?
3. ¿Cuál sería el pronóstico de demanda de bicicletas en cada hora del día dadas unas condiciones?

Supuestos

- 1 El único ingreso de la empresa es por renta de bicicletas
- 2 No hay inversión inicial. Todas las bicicletas y demás equipo para la operación esta disponible.
- 3 Los costos y rentabilidad esperada son parámetros que debe ingresar el usuario
- 4 Cuando el sistema no esta funcionando es porque está en mantenimiento.

LIMPIEZA Y ALISTAMIENTO DE DATOS

Perfilamiento de datos para decidir una estrategia.

Dataset statistics

Number of variables	14
Number of observations	8760
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	2.8 MiB
Average record size in memory	336.8 B

Variable types

DateTime	1
Numeric	10
Categorical	2
Boolean	1

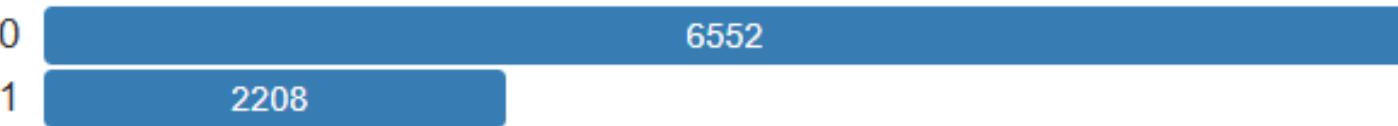
Seasons    
Functioning day (True / False)
Holiday (Holiday / No Holiday)

LIMPIEZA Y ALISTAMIENTO DE DATOS

Seasons_Spring

Categorical

Distinct	2
Distinct (%)	< 0.1%

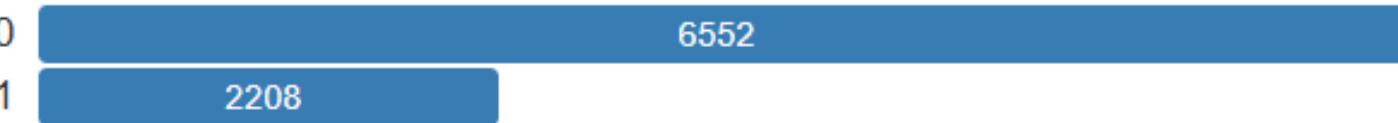


Seasons_Summer

Categorical

HIGH CORRELATION

Distinct	2
Distinct (%)	< 0.1%

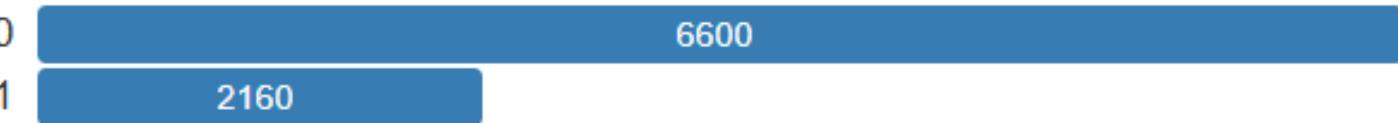


Seasons_Winter

Categorical

HIGH CORRELATION

Distinct	2
Distinct (%)	< 0.1%



LIMPIEZA Y ALISTAMIENTO DE DATOS

Holiday

Categorical

IMBALANCE

Distinct

2

Distinct (%)

< 0.1%



Functioning Day

Categorical

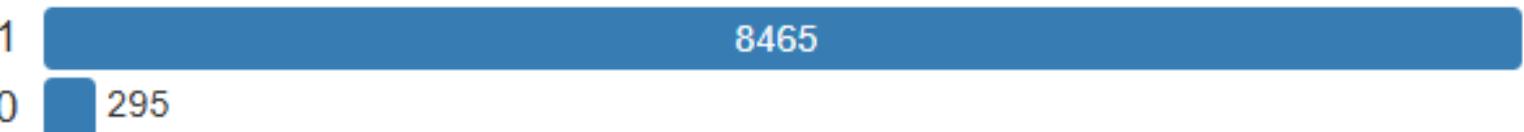
IMBALANCE

Distinct

2

Distinct (%)

< 0.1%



EXPLORACIÓN DE DATOS

En esta sección se realizó un análisis exploratorio de datos para describir estadísticamente y visualizar el comportamiento de las diferentes variables presentes en el dataset.

- Estadísticas descriptivas
- Histogramas
- Diagramas de caja
- Diagramas de dispersión
- Diagramas de violín



ESTADISTICAS DESCRIPTIVAS



Rented bike count

Promedio de 704.6, máximo de 3556, desvío estándar de 644.9.

Temperature

Promedio de 12.88°C, mínimo de -17.8°C, máximo de 39.4°C.

Humidity

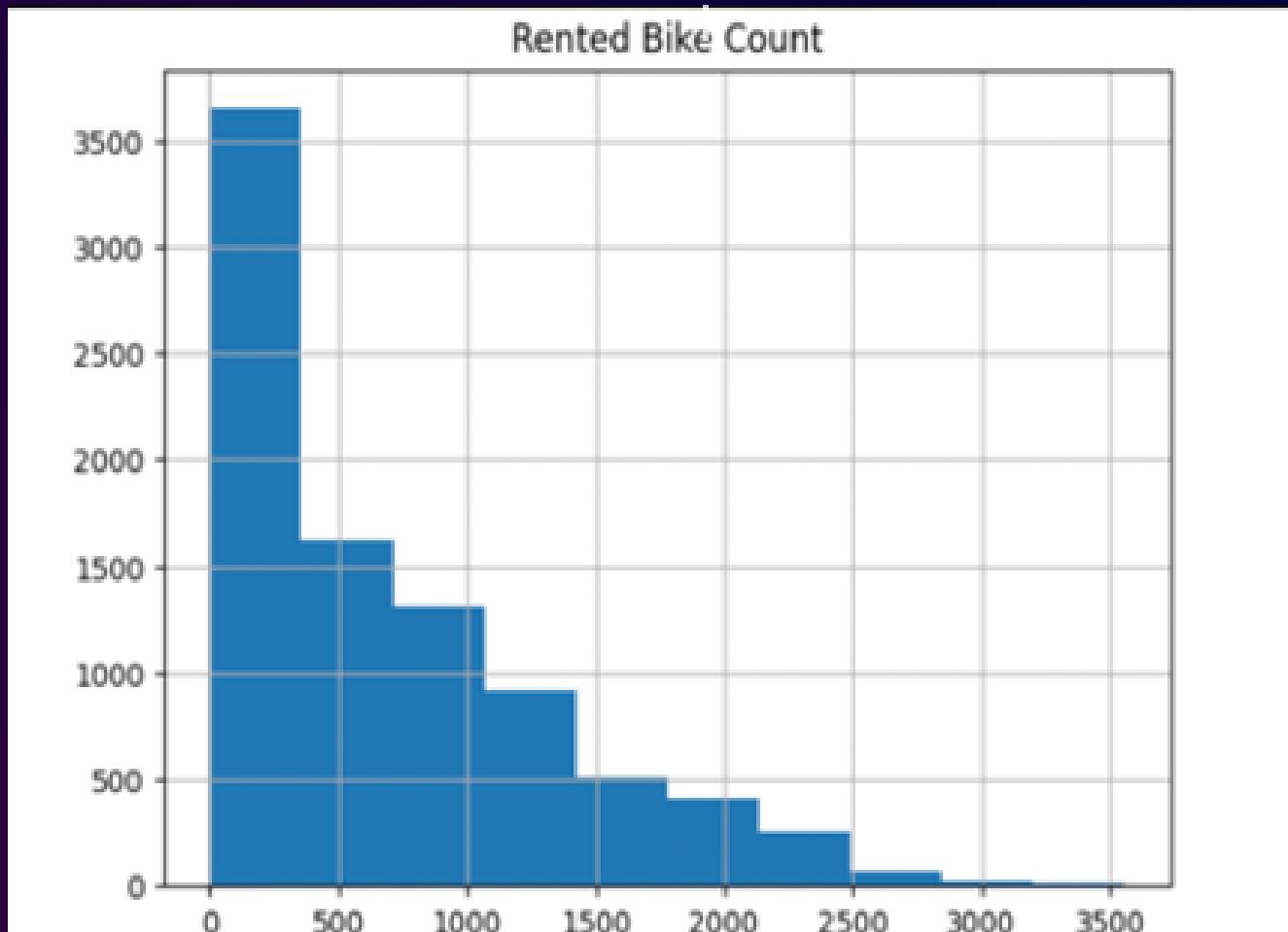
Promedio de 58.22%, variabilidad notable con un mínimo de 0% y un máximo de 98%.

Wind speed

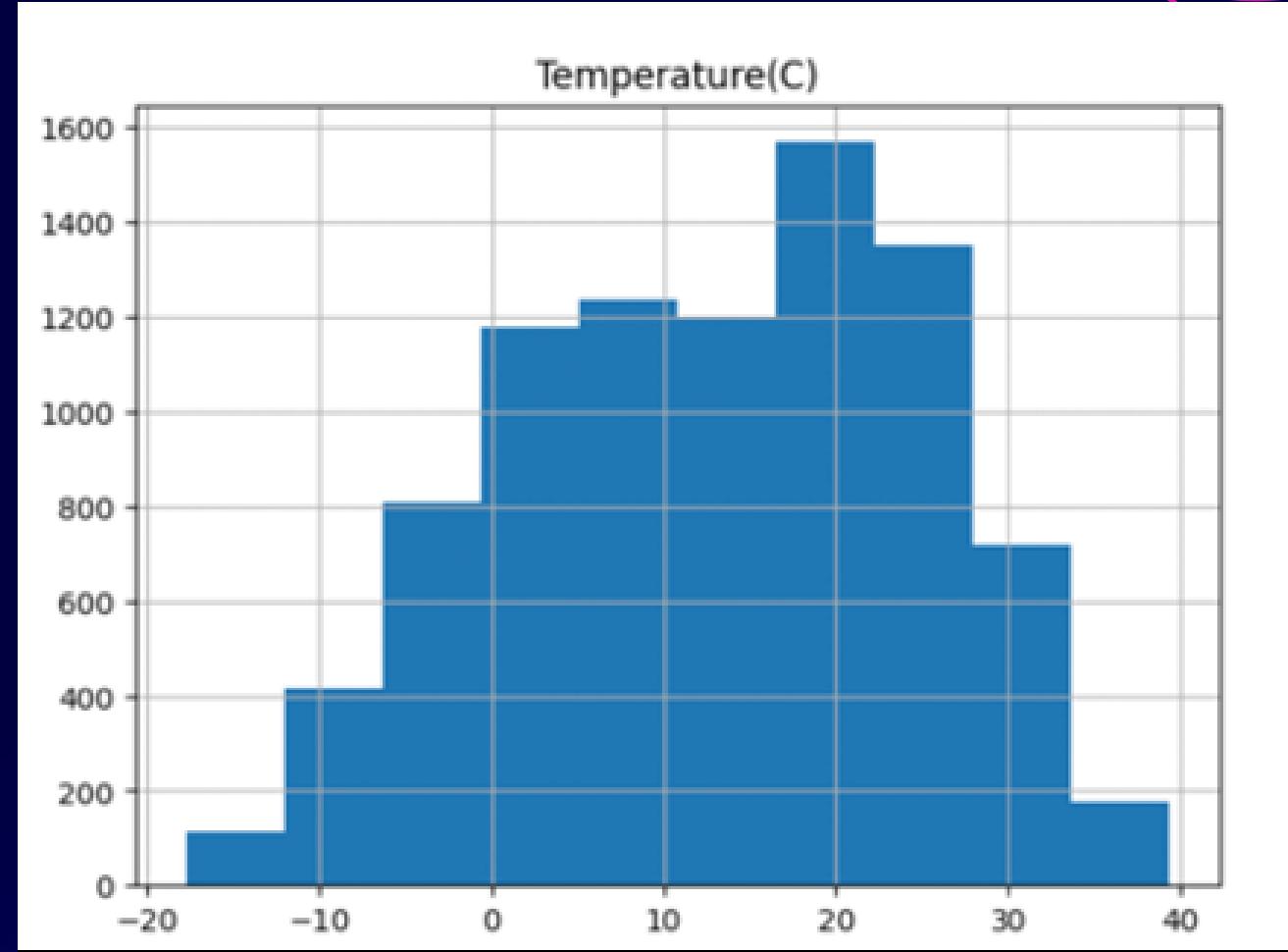
Velocidad media de 1.72 m/s, con una variabilidad de hasta 7.4 m/s.



HISTOGRAMAS

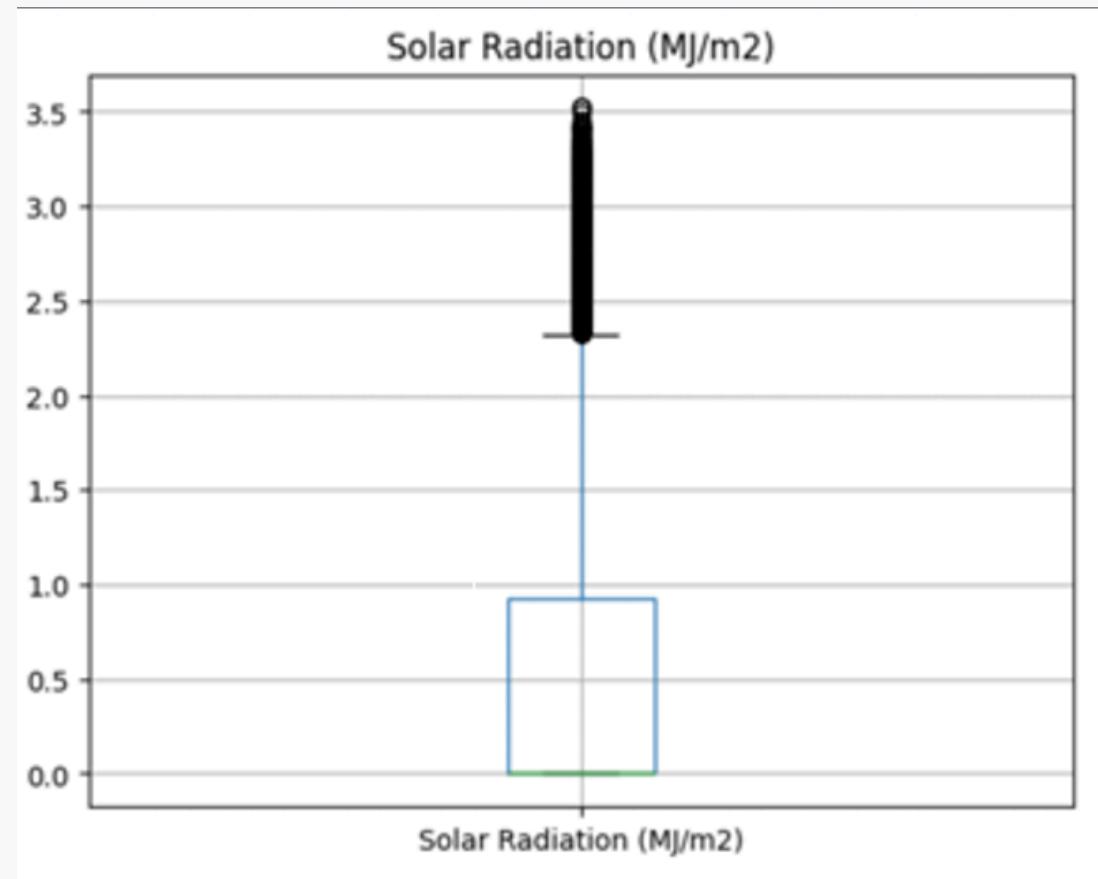


La mayoría de veces en el día existe una baja demanda de bicicletas rentadas, lo que indica que una alta demanda puede depender de factores externos específicos.

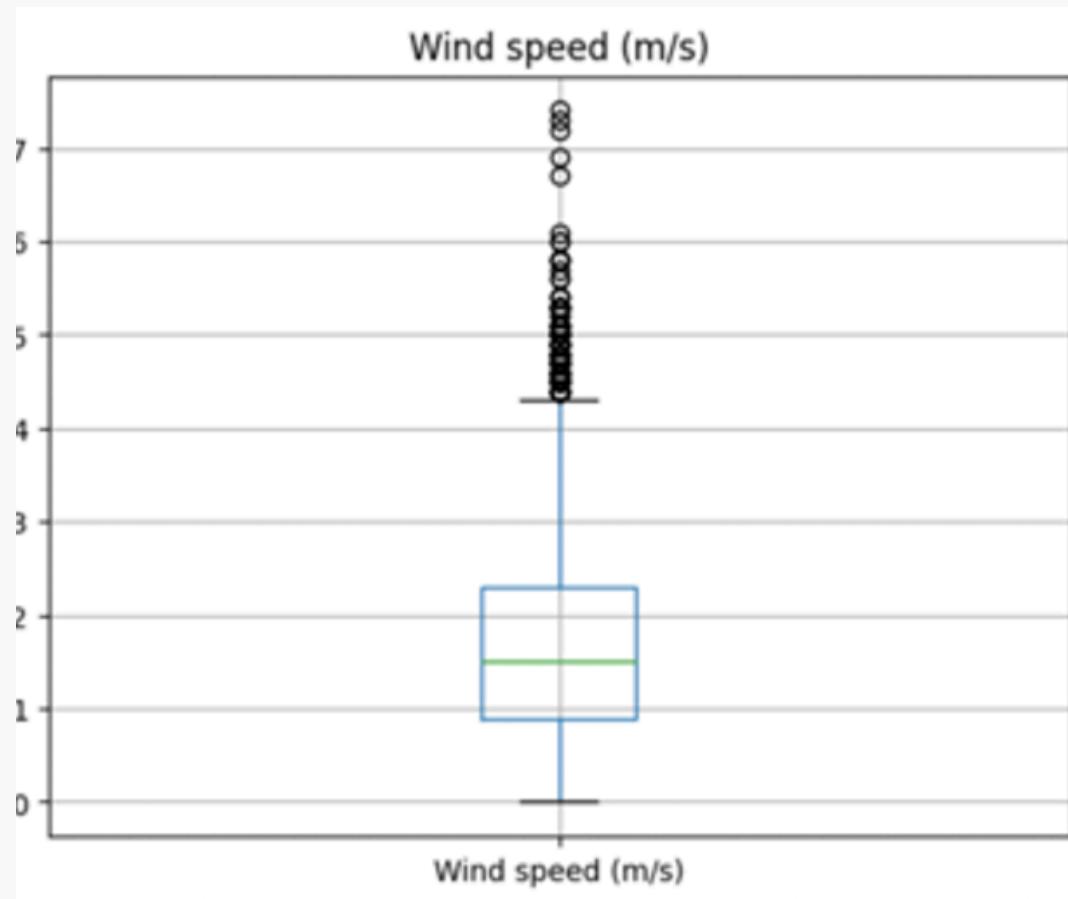


Variables como la temperatura, que sugieren que existen climas extremos en la región de estudio, probablemente pueden afectar negativamente la cantidad de bicicletas rentadas. Puesto que, situaciones climáticas extremas puede afectar la cantidad de gente que sale a las calles.

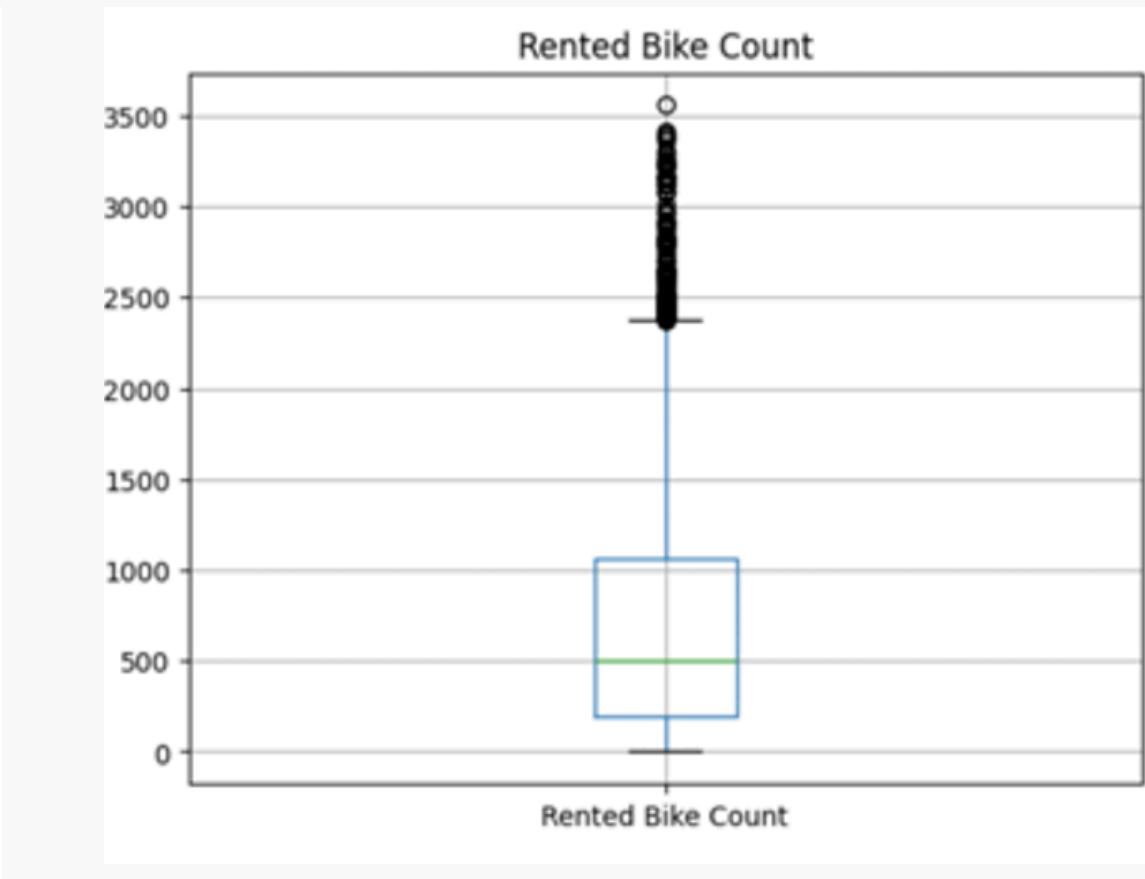
BOX PLOT



La radiación solar tiene una alta concentración de valores atípicos hacia el extremo superior. Esto indica que aunque la mayoría de los días tienen baja radiación solar, hay ciertos días con radiación extremadamente alta.

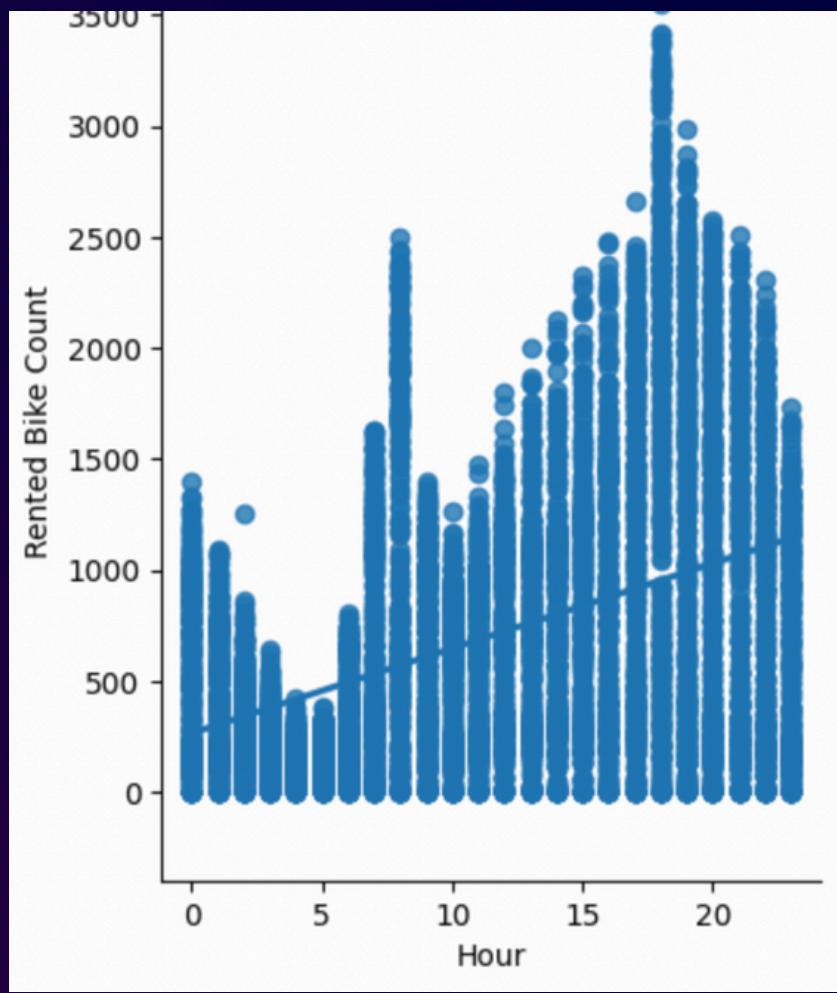


La velocidad del viento presenta valores atípicos por encima del límite superior, lo que indica que días con vientos fuertes son excepcionales. Esto es importante, ya que puede tener implicaciones para la seguridad y comodidad de los ciclistas, afectando la demanda en esos días.

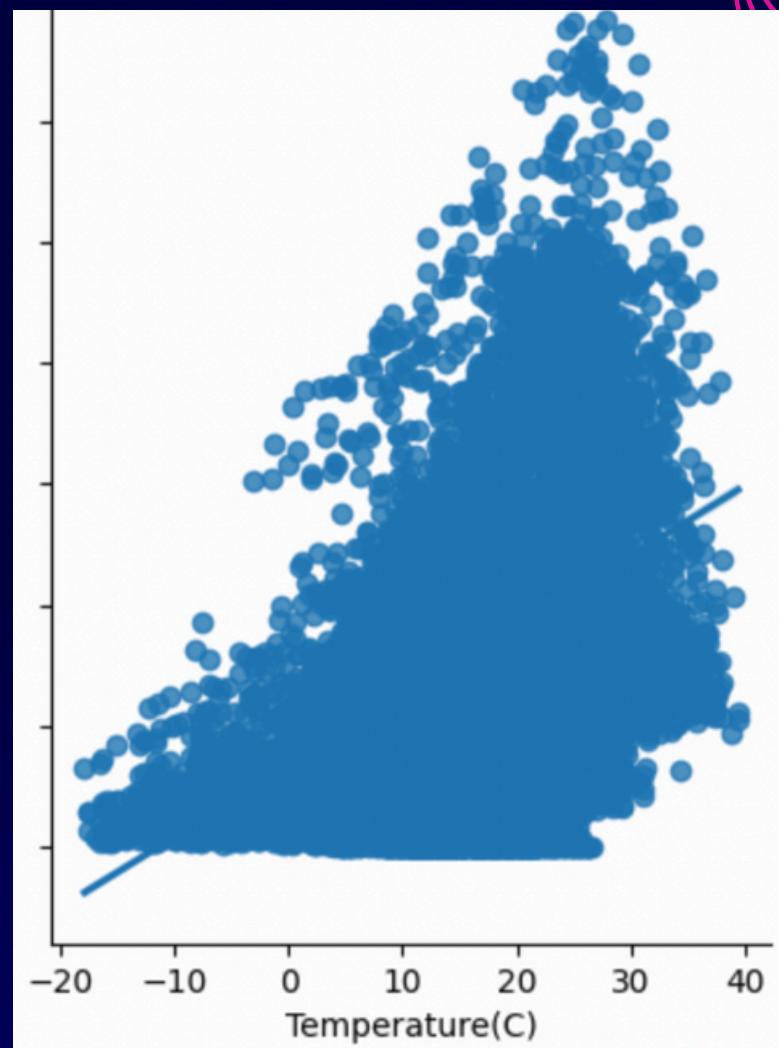


Gran cantidad de valores atípicos por encima del límite superior. Esto confirma la naturaleza esporádica de la alta demanda observada anteriormente y sugiere que ciertas condiciones o eventos pueden provocar picos en el uso.

DIAGRAMAS DE DISPERSIÓN

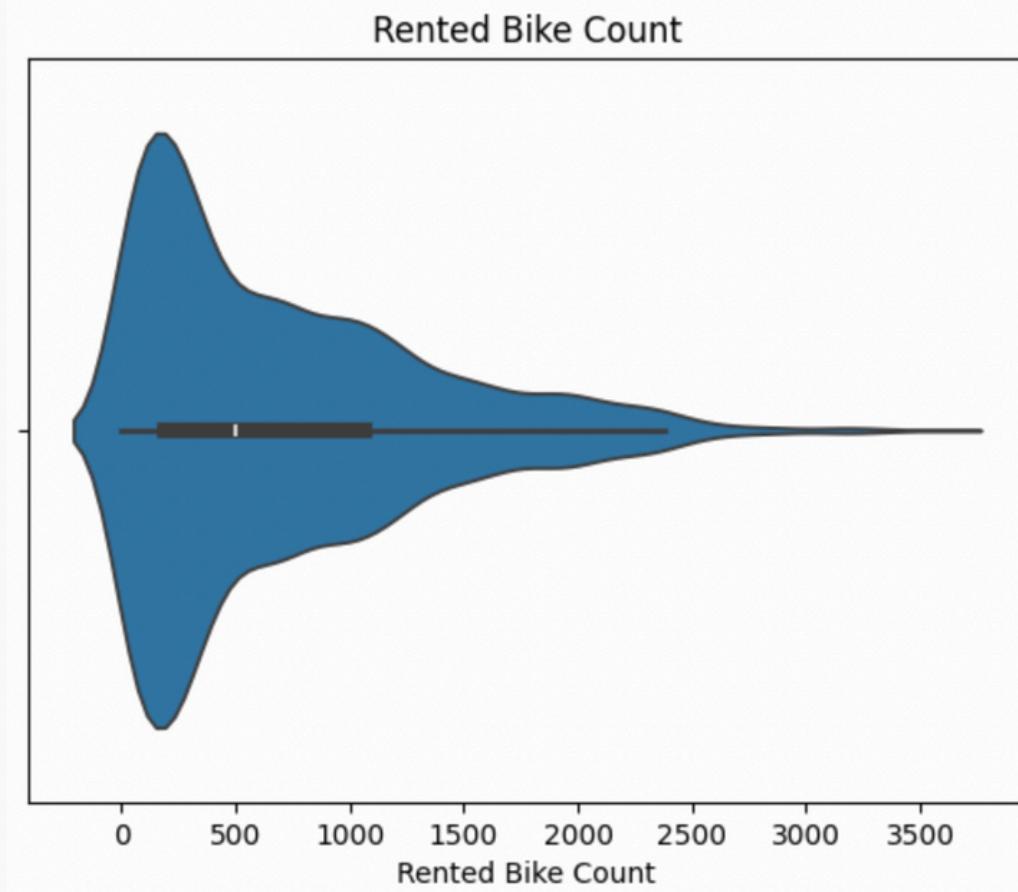


Existe una relación positiva clara entre la hora del día y el número de bicicletas rentadas, alcanzando un pico en la tarde (alrededor de las 17:00-18:00), lo que indica una alta demanda posiblemente debido a los desplazamientos laborales.

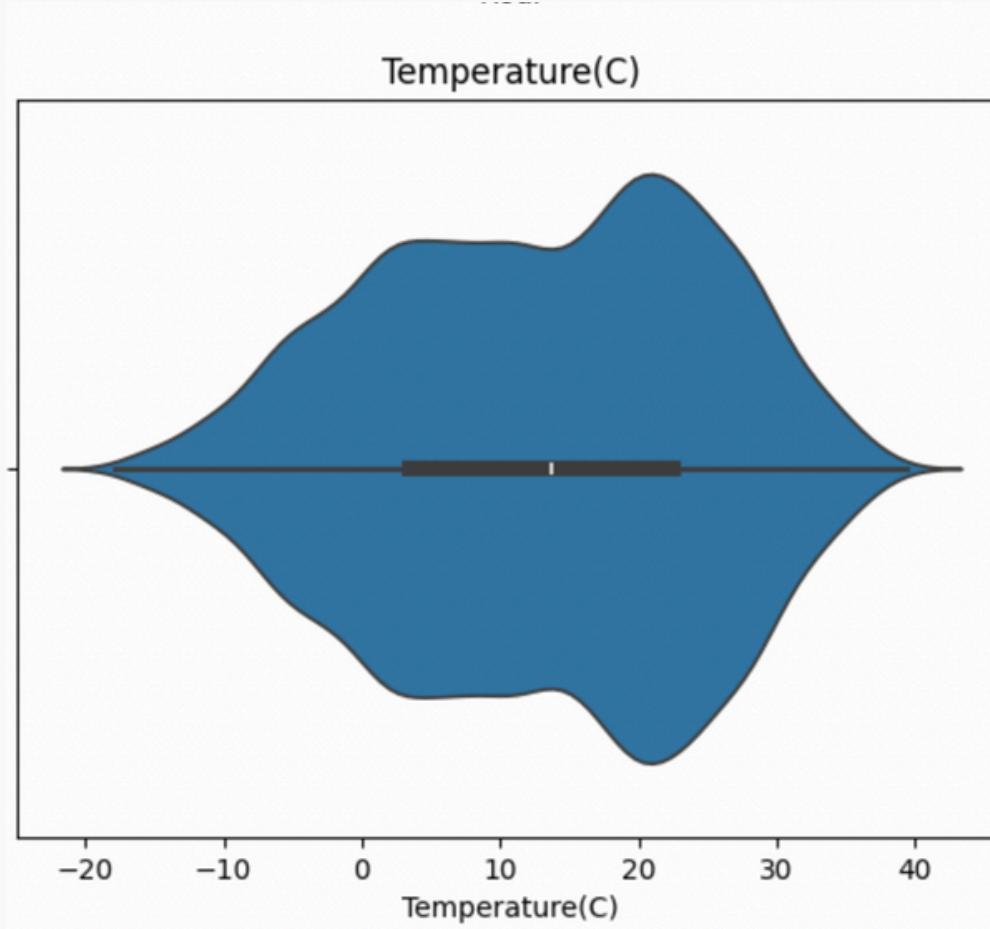


La demanda de bicicletas aumenta con la temperatura hasta cierto punto (alrededor de 20-25°C), lo que sugiere que condiciones climáticas moderadas son ideales para el uso de bicicletas.

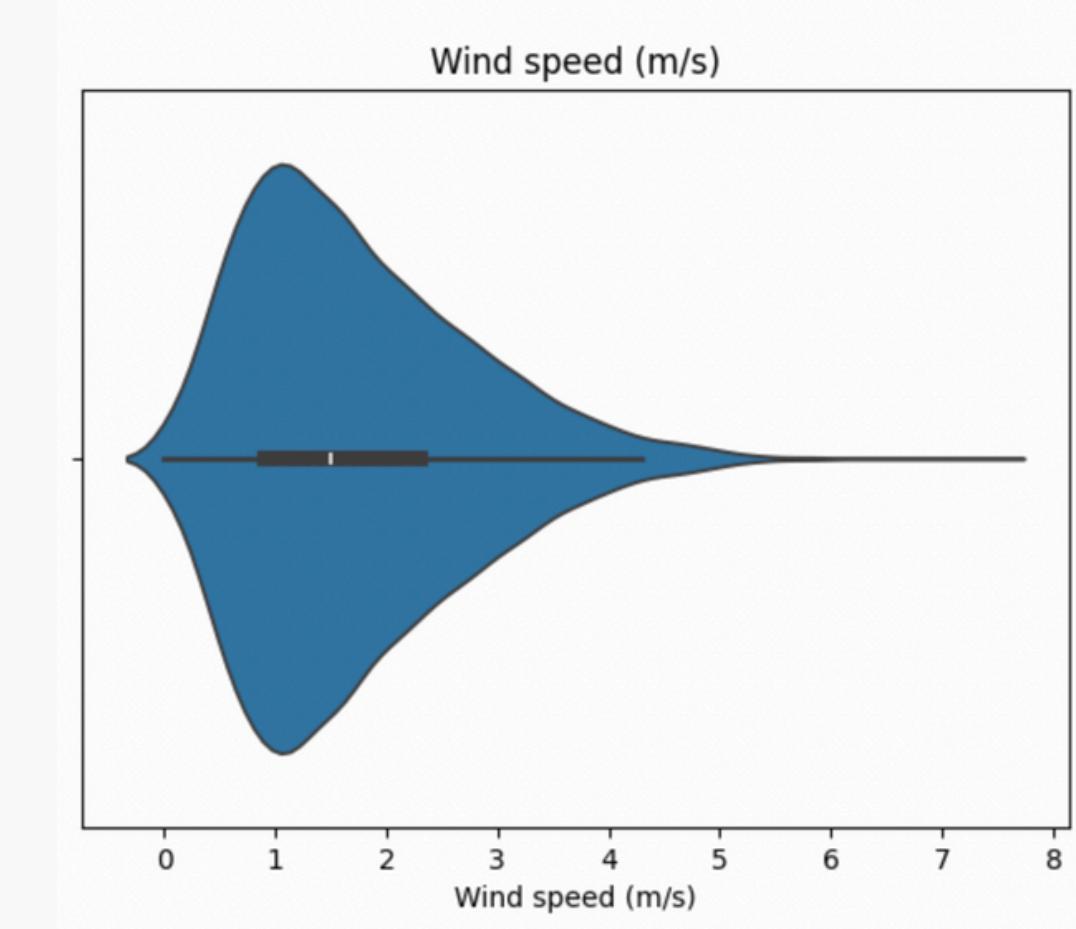
DIAGRAMAS DE VIOLIN



La distribución muestra una asimetría a la izquierda con una alta densidad de datos en valores bajos de rentas de bicicletas. Esto sugiere que la mayoría de los conteos de bicicletas rentadas son bajos, con algunas observaciones que representan un conteo mucho mayor.



La distribución de la temperatura es aproximadamente simétrica alrededor de la media, con una mayor concentración de datos entre 10°C y 30°C



La distribución de la velocidad del viento tiene una alta concentración alrededor de valores bajos (cerca de 1-2 m/s) y disminuye rápidamente a medida que la velocidad del viento aumenta.

MODELAMIENTO

- Se desarrolló un modelo de regresión lineal con 14 variables para estimar la cantidad de bicicletas rentadas por hora (Rented bike count).
- La regresión lineal, según Rumsey (2016), permite entender y predecir una variable dependiente basada en variables independientes.
- Tras el análisis estadístico, se eliminaron variables no significativas, logrando una estimación más precisa.
- Se presentan los modelos generados y el modelo final seleccionado para interpretar la cantidad de bicicletas rentadas.



MODELO INCIAL



$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} \\ + \beta_{11} X_{11} + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14}$$

Y: Rented Bike Count

X1: Hour

X2: Temperature(C)

X3: Humidity(%)

X4: Wind speed (m/s)

X5: Visibility (10m)

X6: Dew point temperature(C)

X7: Solar Radiation (MJ/m2)

X8: Rainfall(mm)

X9: Snowfall (cm)

X10: Holiday

X11: Functioning Day

X12: Seasons_Spring

X13: Seasons_Summer

X14: Seasons_Winter

B0: 53.98370399970281

B1 (Hour): 27.36755343612323

B2 (Temperature(C)): 14.43826753216662

B3 (Humidity(%)): -11.103888587537377

B4 (Wind speed (m/s)): 19.496121345722678

B5 (Visibility (10m)): 0.01745474269429792

B6 (Dew point temperature(C)): 12.820752006513155

B7 (Solar Radiation (MJ/m2)): -75.87692322007723

B8 (Rainfall(mm)): -55.83998938115196

B9 (Snowfall (cm)): 29.026841185932145

B10 (Holiday): -115.79557975811028

B11 (Functioning Day): 931.2135047213319

B12 (Seasons_Spring): -125.83055751023507

B13 (Seasons_Summer): -142.94948619323395

B14 (Seasons_Winter): -357.4897874909983

MODELO INICIAL

- Para el modelo de regresión lineal se hace un entrenamiento con una división de los datos del 20-80. Esto con el fin de entrenar el modelo y después probar con los mismos datos que tan bueno es el ajuste para estimar el comportamiento de la variable dependiente. Esto es necesario para evitar sesgos de evaluación y obtener resultados más fiables y generalizables (Sangha, 2021)
- Revisión de las medidas de desempeño:

MAE: 326.4871946346789

MSE: 190609.26508288088

RMSE: 436.58820996779207

R2: 0.5445663619093557

- El modelo es significativo. (F-statistic=0,00)
- La variable visibilidad no es significativa (P-value=0,258)



MODELO FINAL

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} \\ + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14}$$

- Revisión de las medidas de desempeño:

MAE: 326.60015025036734

MSE: 190408.074931489

RMSE: 436.357737334276

R2: 0.5450470770653402

- El modelo es significativo.
- Todas las variables del modelo son significativas.



DISTRIBUCION DEL TABLERO

Datos Historicos

Histograma de la distribución de la demanda de renta de bicicletas a través del tiempo para cada estación.

Predicción de demanda

En esta sección se estima la demanda de renta de bicicletas a través del modelo de regresión lineal explicado anteriormente.

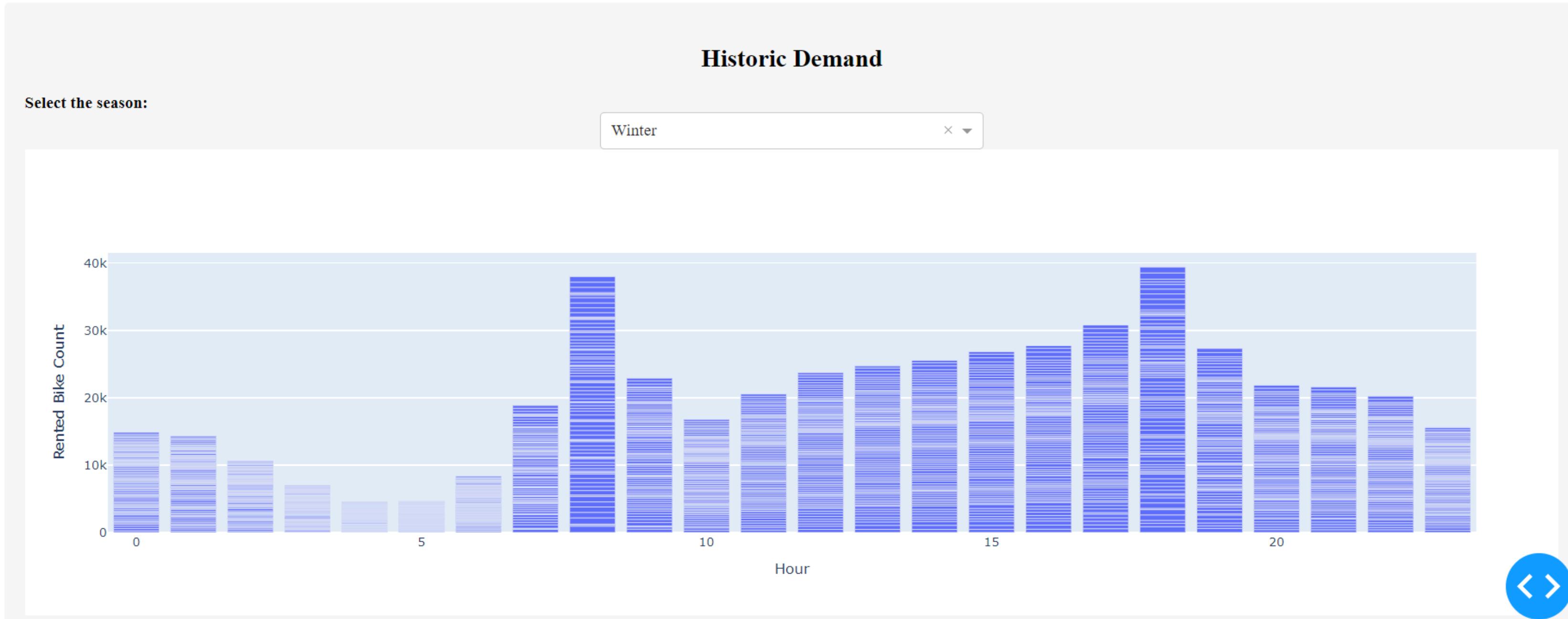
Estimación del precio

Finalmente con la demanda estimada y con los costos como insumo (parámetros) se estima el precio que debe tener una bicicleta para cubrir costos y llegar a rentabilidad.

$$precio = \frac{Costo Fijo + Costo Variable * Demanda}{Demanda} + Rentabilidad Esperada$$

PRODUCTO FINAL

Seoul Bike Sharing Demand



PRODUCTO FINAL

Demand Prediction

Select variables for demand prediction:

Holiday
 Functioning Day

Select the season:

Spring

Temperature:



Rainfall:



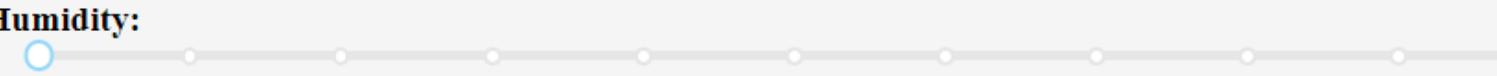
Solar Radiation:



Snowfall:



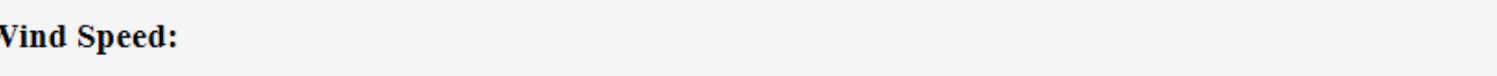
Humidity:



Hour:



Wind Speed:



Dew Point Temperature:



Predict

Demand Predicted:

CI: (286.24 , 1541.97) Media: 914.11



PRODUCTO FINAL

Cost Analysis

Enter fixed cost per hour:

Enter variable cost per bike:

Enter profitability (decimal):

Calculate

Suggested Price per Bike per hour:

7.21

Hour Expenses

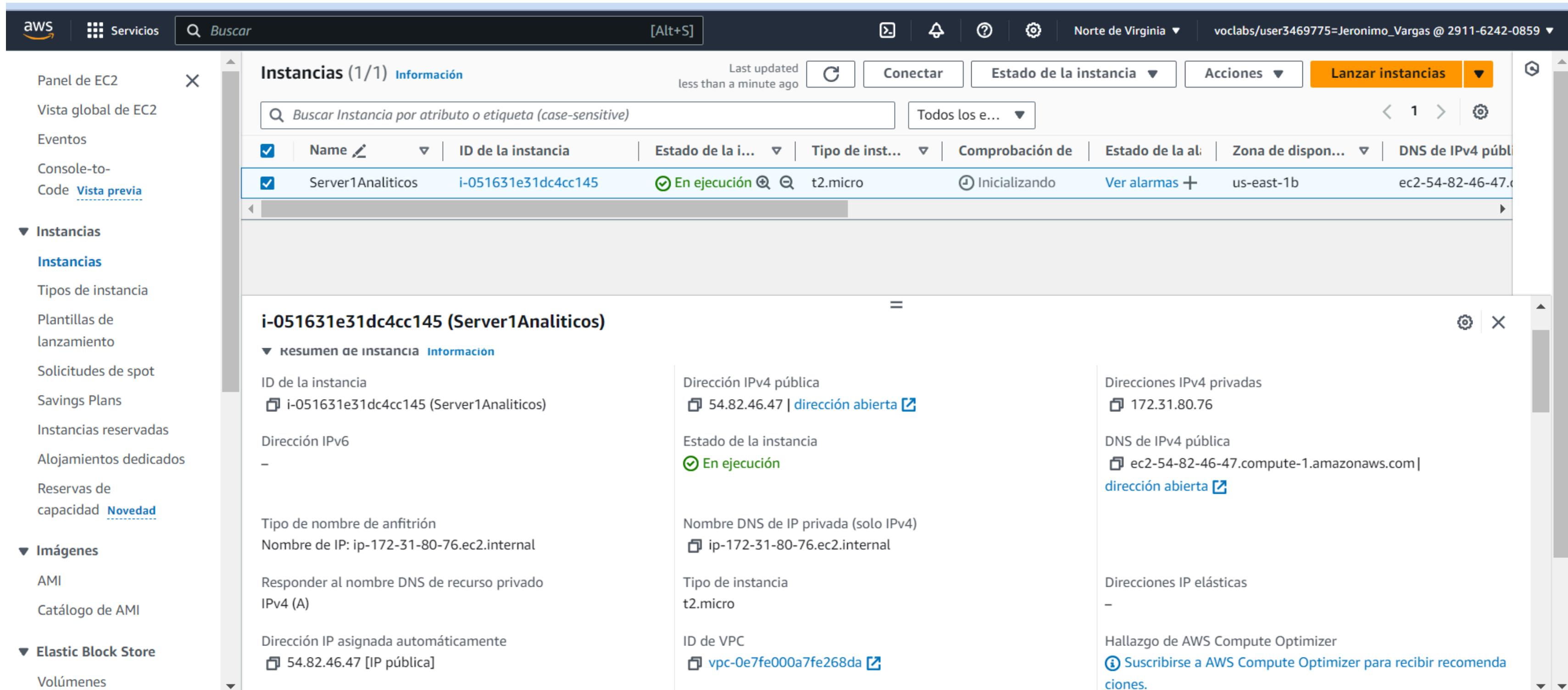
Cost Type	Percentage
Variable Cost	99.9%
Fixed Cost	0.0781%

Variable Cost
Fixed Cost

<>

DESPLIEGUE EC2

Tipo de instancia: t2.micro
Plataforma y Detalles de la plataforma: Amazon Linux (inferido), Linux/UNIX
Tipo de virtualización: hvm
Número de CPU virtuales: 1
Zona de disponibilidad (lo encuentra en la pestaña Redes): us-east-1b



The screenshot shows the AWS EC2 Instances page. On the left, there's a sidebar with navigation links like 'Panel de EC2', 'Vista global de EC2', 'Eventos', 'Console-to-Code', and 'Instancias'. Under 'Instancias', there are sub-links for 'Instancias', 'Tipos de instancia', 'Plantillas de lanzamiento', 'Solicitudes de spot', 'Savings Plans', 'Instancias reservadas', 'Alojamientos dedicados', 'Reservas de capacidad', 'Imagenes', 'AMI', 'Catálogo de AMI', and 'Elastic Block Store', 'Volúmenes'. The main content area shows 'Instancias (1/1) Información'. It lists one instance: 'Server1Analiticos' (ID: i-051631e31dc4cc145). The instance is in 'En ejecución' state, type 't2.micro', and is associated with a VPC interface (vpc-0e7fe000a7fe268da). It has a public IP of 54.82.46.47 and a private DNS name of ip-172-31-80-76.ec2.internal. The 'Resumen de instancia' section provides detailed information about the instance's configuration, including its name, ID, and network details.

Name	ID de la instancia	Estado de la i...	Tipo de inst...	Comprobación de	Estado de la al...	Zona de dispon...	DNS de IPv4 públ...
Server1Analiticos	i-051631e31dc4cc145	En ejecución	t2.micro	Inicializando	Ver alarmas	us-east-1b	ec2-54-82-46-47...

i-051631e31dc4cc145 (Server1Analiticos)

Resumen de instancia

ID de la instancia	Dirección IPv4 pública	Direcciones IPv4 privadas
i-051631e31dc4cc145 (Server1Analiticos)	54.82.46.47 dirección abierta	172.31.80.76
Dirección IPv6	Estado de la instancia	DNS de IPv4 pública
-	En ejecución	ec2-54-82-46-47.compute-1.amazonaws.com dirección abierta
Tipo de nombre de anfitrión	Nombre DNS de IP privada (solo IPv4)	Direcciones IP elásticas
Nombre de IP: ip-172-31-80-76.ec2.internal	ip-172-31-80-76.ec2.internal	-
Responder al nombre DNS de recurso privado	Tipo de instancia	Hallazgo de AWS Compute Optimizer
IPv4 (A)	t2.micro	Suscribirse a AWS Compute Optimizer para recibir recomendaciones.
Dirección IP asignada automáticamente	ID de VPC	
54.82.46.47 [IP pública]	vpc-0e7fe000a7fe268da	

DESPLEGUE EC2

```
ec2-user@ip-172-31-80-76:~ x Windows PowerShell x + v
C:\Users\Jeronimo Vargas\.ssh/known_hosts:6: 35.174.115.199
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added '54.82.46.47' (ED25519) to the list of known hosts.

A newer release of "Amazon Linux" is available.
Version 2023.5.20240903:
Run "/usr/bin/dnf check-release-update" for full release and version update info
,
#_
~\_\_ #####_ Amazon Linux 2023
~~ \_\#\#\#\#\_
~~ \#\#\#
~~ \#/ ___ https://aws.amazon.com/linux/amazon-linux-2023
~~ V~' '-->
~~~ /
~~ ._. /_
~/ _/ /
~/m/'

Last login: Mon Sep  9 18:02:35 2024 from 201.244.245.206
[ec2-user@ip-172-31-80-76 ~]$ python3 appdahs.py
Archivo SeoulBikeDataClean.csv cargado desde: SeoulBikeDataClean.csv
Archivo SeoulBikeData_utf8.csv cargado desde: SeoulBikeData_utf8.csv
Dash is running on http://0.0.0.0:8050/

* Serving Flask app 'appdahs'
* Debug mode: on
Archivo SeoulBikeDataClean.csv cargado desde: SeoulBikeDataClean.csv
Archivo SeoulBikeData_utf8.csv cargado desde: SeoulBikeData_utf8.csv
201.244.245.206 - - [10/Sep/2024 01:48:08] code 400, message Bad request version ('\\x00\\x12\\x00\\x10\\x04\\x03\\x08\\x04\\x01\\x05\\x03\\x08\\x05\\x01\\x08\\x06\\x06\\x01\\x01\\x00\\x01\\x00\\x00')
201.244.245.206 - - [10/Sep/2024 01:48:08] code 400, message Bad request syntax ('\\x16\\x03\\x01\\x06à\\x01\\x00\\x06Ü\\x03\\x03')
201.244.245.206 - - [10/Sep/2024 01:48:08] code 400, message Bad request version ('^øÌYý')
201.244.245.206 - - [10/Sep/2024 01:48:08] code 400, message Bad request version ("\\x8eT¤r¥>Bü'%5\\xabb\\x80/$\\x12ÍÖ\\xad¥ü^\\x94ñ\\x9d7é')
```